Nonstationary Panels

Based on chapters 12.4, 12.5, and 12.6 of Baltagi, B. (2005): Econometric Analysis of Panel Data, 3rd edition. Chichester, John Wiley & Sons.

12.4 Spurious Regressions in Panel Data

- ✓ Entorf (1997) studied spurious fixed effects regressions when the true model involves independent random walks with and without drifts.
- ✓ Phillips and Moon provided a regression limit theory for non stationary panel data with large numbers of cross section and time series.
- ✓ Kao studied the Least-Squares Dummy Variable estimator (LSDV) where the spurious regression phenomenon is still present for independent nonstationary variables.

- Entorf (1997): found that for $T \rightarrow \infty$ and *N* finite the nonsense regression phenomenon holds for spurious fixed effects models and inference based on *t*-values can be highly misleading.
- This implies seemingly significant t-statistics and high R^2 in case of FE estimation

- Suppose:
  - y_t *and* X_t *are unit root nonstationary time series variables*
  - with long-run variance matrix

$$\Omega = \begin{pmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{pmatrix}$$

  - Then $\beta = \Omega_{yx}\Omega_{xx}^{-1}$ can be interpreted as a classical long-run regression coefficient relating *the two nonstationary variables* y_t *and* $X_{t.}$
  - When Ω has deficient rank, β is a cointegrating coefficient because $y_t - \beta X_t$ *is stationary.*

- Phillips and Moon (1999) extend this concept to panel regressions with nonstationary data.
- In this case, heterogeneity across individuals *i* can be characterized by heterogeneous long-run covariance matrices Ω_i .
- Then Ω_i are randomly drawn from a population with mean $\Omega = E(\Omega i)$.
- In this case:
  - the regression coefficient corresponding to the average long-run covariance matrix is

$$\beta = E[\Omega_{y_i x_i}] E[\Omega_{x_i x_i}]^{-1} = \Omega_{yx} \Omega_{xx}^{-1}$$

Hence, we get a fundamental framework for studying sequential and joint limit theories in nonstationary panel data, which allows for four cases:

1. Panel spurious regression
2. Heterogeneous panel cointegration
3. Homogeneous panel cointegration
4. Near-homogeneous panel cointegration

- Phillips and Moon (1999) investigated these four models and developed panel asymptotics for regression coefficients and tests using both sequential and joint limit arguments.
- In all four cases
  - The pooled estimator is consistent and has a normal limiting distribution.
  - The pooled least squares estimator of the slope coefficient β is √*N-consistent for the long-run average relation parameter* β *and has a limiting normal distribution.*
  - A limiting cross-section regression with time-averaged data is also √*N-consistent for* β *and* has a limiting normal distribution.

- This is different from the pure time series spurious regression where the limit of the OLS estimator of β is a nondegenerate random variate that is a functional of Brownian motions and is therefore not consistent for β.
- The idea in Phillips and Moon (1999) is that independent cross-section data in the panel adds information and this leads to a stronger overall signal than the pure time series case.

12.5 PANEL COINTEGRATION TESTS

- Like the panel unit root tests, panel cointegration tests can be motivated by the search for more powerful tests than those obtained by applying individual time series cointegration tests.
- In the case of purchasing power parity and convergence in growth, economists pool data on similar countries, like G7, OECD or Euro countries in the hopes of adding cross-sectional variation to the data that will increase the power of unit root tests or panel cointegration tests.

- Null of no cointegration
  - **Residual-Based DF and ADF Tests (Kao Tests)**
- Null of cointegration
  - **Residual-Based LM Test**
  - **Pedroni Tests**

Residual-Based DF and ADF Tests (Kao Tests)

- the panel regression model:

$$y_{it} = x_{it}'\beta + z_{it}'\gamma + e_{it}$$

Where y_{it} and x_{it} are I(1) and noncointegrated.

- For $z_{it} = \{\mu_i\}$, Kao(1999) proposed Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) type unit root tests for e_{it} as a test for the null of no cointegration.
- The DF-type tests can be calculated from the fixed effects residuals

$$\hat{e}_{it} = \rho\hat{e}_{it-1} + \nu_{it}$$

$$\text{where } \hat{e}_{it} = \tilde{y}_{it} - \tilde{x}_{it}'\beta, \tilde{y}_{it} = y_{it} - \bar{y}_{i\cdot}$$

- To test the null hypothesis of no cointegration

$$H_0 : \rho = 1$$

- The OLS estimate of ρ and the *t-statistic are given as*

$$\hat{\rho} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{e}_{it}\hat{e}_{it-1}}{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{e}_{it}^2}$$

$$t_{\hat{\rho}} = \frac{(\hat{\rho}-1)\sqrt{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{e}_{it-1}^2}}{s_e}$$

where $s_e^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{e}_{it} - \hat{\rho}\hat{e}_{it-1})^2$

- Kao proposed the following four DF-type tests:

$$DF_\rho = \frac{\sqrt{N}T(\hat{\rho}-1)+3\sqrt{N}}{\sqrt{10.2}}$$

$$DF_t = \sqrt{1.25}t_{\hat{\rho}} + \sqrt{1.875N}$$

$$DF_\rho^* = \frac{\sqrt{N}T(\hat{\rho}-1)+\frac{3\sqrt{N}\hat{\sigma}_\nu^2}{\hat{\sigma}_{0\nu}^2}}{\sqrt{3+\frac{36\hat{\sigma}_\nu^4}{5\hat{\sigma}_{0\nu}^4}}}$$

$$DF_t^* = \frac{t_{\hat{\rho}}+\frac{\sqrt{6N}\hat{\sigma}_\nu}{2\hat{\sigma}_{0\nu}}}{\sqrt{\frac{\hat{\sigma}_{0\nu}^2}{2\hat{\sigma}_\nu^2}+\frac{3\hat{\sigma}_\nu^2}{10\hat{\sigma}_{0\nu}^2}}}$$

where $\hat{\sigma}_\nu^2 = \hat{\Sigma}_{yy} - \hat{\Sigma}_{yx}\hat{\Sigma}_{xx}^{-1}$, $\hat{\sigma}_{0\nu}^2 = \hat{\Omega}_{yy} - \hat{\Omega}_{yx}\hat{\Omega}_{xx}^{-1}$

- While DF_ρ and DF_t are based on the strong exogeneity of the regressors and errors, $DF_{*\rho}$ and DF_{*t} are for the cointegration with endogenous relationship between regressors and errors.
- For the ADF test, we can run the following regression:

$$\hat{e}_{it} = \rho \hat{e}_{it-1} + \sum_{j=1}^{p} \vartheta_j \Delta \hat{e}_{it-j} + \nu_{itp}$$

- With the null hypothesis of no cointegration,

$$ADF = \frac{t_{ADF} + \frac{\sqrt{6N}\hat{\sigma}_\nu}{2\hat{\sigma}_{0\nu}}}{\sqrt{\frac{\hat{\sigma}_{0\nu}^2}{2\hat{\sigma}_\nu^2} + \frac{3\hat{\sigma}_\nu^2}{10\hat{\sigma}_{0\nu}^2}}}$$

Residual-Based LM Test

- McCoskey and Kao (1998) derived a residual-based test for the null of cointegration rather than the null of no cointegration in panels.
- This test is an extension of the LM test and the locally best invariant (LBI) test for an MA unit root in the time series literature.
- Under the null, the asymptotics no longer depend on the asymptotic properties of the estimating spurious regression, rather the asymptotics of the estimation of a cointegrated relationship are needed.

Pedroni Tests

- Pedroni (2000, 2004) also proposed several tests for the null hypothesis of cointegration in a panel data model that allows for considerable heterogeneity.
- His tests can be classified into two categories.
  - The first set is similar to the tests discussed above, and involves averaging test statistics for cointegration in the time series across cross-sections.
  - For the second set, the averaging is done in pieces so that the limiting distributions are based on limits of piecewise numerator and denominator terms.

Finite Sample Properties

- McCoskey and Kao (1999) conducted Monte Carlo experiments to compare the size and power of different residual-based tests for cointegration in heterogeneous panel data: varying slopes and varying intercepts.
- They found that the average ADF performs better with respect to power and their maximum eigenvalue-based p-value performs better with regard to size.
- The test of the null hypothesis was originally proposed in response to the low power of the tests of the null of no cointegration, especially in the time series case.
- Authors find that in cases, where economic theory predicts long-run steady-state relationships, it seemed that a test of the null of cointegration rather than the null of no cointegration would be appropriate.

12.6 ESTIMATION AND INFERENCE IN PANEL COINTEGRATION MODELS

- The asymptotic properties of the estimators of the regression coefficients and the associated statistical tests are different from those of the time series cointegration regression models.
- The finite sample proprieties of the OLS estimator the t-statistic, the bias-corrected OLS estimator, and the bias-corrected t-statistic.
- The bias-corrected OLS estimator does not improve over the OLS estimator in general.

- Kao and Chiang (2000) consider the following panel regression:

$$y_{it} = x_{it}'\beta + z_{it}'\gamma + u_{it}$$

Where

$$x_{it} = x_{it-1} + \varepsilon_t$$

- *y_{it} is cointegrated with x_{it}*
- The assumption of cross-sectional independence is maintained.
- The OLS estimator of β is

$$\hat{\beta}_{OLS} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{y}_{it}\right)$$

- β_{OLS} is inconsistent using panel data.
- This is in sharp contrast with the consistency of β_{OLS} in time series under similar circumstances.

- Choi (2002) studied instrumental variable estimation for an error component model with stationary and nearly nonstationary regressors.

- Consider the simple panel regression

$$y_{it} = \alpha + \beta x_{it} + u_{it}$$

- Where x_{it} is nearly nonstationary, u_{it} is *I*(0) and z_t is an instrumental variable
- Yielding the panel IV (Within) estimator

$$\hat{\beta}_{IV} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_{i\cdot})(z_{it} - \bar{z}_{i\cdot})\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_{i\cdot})(z_{it} - \bar{z}_{i\cdot})\right]$$

- when *N is large, and proper conditions hold, the central limit theorem can be applied which leads to the asymptotic normality result for the panel estimator.*